

STATA を用いた臨床試験解析

STATA にて多発性骨髄腫の臨床試験データを解析することにしました。まず STATA を立ち上げ、Editor Box を開き下記データを入力します。他のソフトからもコピーできます。Var1 等は double click することにより変数に名前をつけることができます。全部記入し終わったら Preserve Box を選び、一端を閉じます。File box より save as を選ぶと*.dta file という文字がでますので、適当に名前をつけて保存する場所を確保しておきましょう。保存には Log Box をクリックすると黒い画面のデータが白い画面に移って保存されます。STATA では黒い画面のデータ自体保存できないので注意してください。さて黒い画面下のコマンドボックスに list と入力してみてください。データが打ち出されます。

```
. edit
```

```
- preserve
```

```
. list
```

	patid	wbc	time	aggroup	Censor1	Censor2
1.	1	2300	65	1	1	1
2.	2	750	156	1	1	1
3.	3	4300	100	1	1	1
4.	4	2600	134	1	1	1
5.	5	6000	16	1	1	0
6.	6	10500	108	1	1	1
7.	7	10000	121	1	1	1
8.	8	17000	4	1	1	1
9.	9	5400	39	1	1	1
10.	10	7000	143	1	1	0
11.	11	9400	56	1	1	1
12.	12	32000	26	1	1	1
13.	13	35000	22	1	1	1
14.	14	100000	1	1	1	1
15.	15	100000	1	1	1	0
16.	16	52000	5	1	1	1
17.	17	100000	65	1	1	1
18.	18	4400	56	2	1	1

19.	19	3000	65	2	1	1
20.	20	4000	17	2	1	0
21.	21	1500	7	2	1	1
22.	22	9000	16	2	1	1
23.	23	5300	22	2	1	1
24.	24	10000	3	2	1	1
25.	25	19000	4	2	1	0
26.	26	27000	2	2	1	1
27.	27	28000	3	2	1	1
28.	28	31000	8	2	1	1
29.	29	26000	4	2	1	1
30.	30	21000	3	2	1	0
31.	31	79000	30	2	1	1
32.	32	100000	4	2	1	1
33.	33	100000	43	2	1	1

ここで、survival command に入る前にこのデータが survival data であることをコンピュータに伝えなくてはなりません。

```
stset time Censor1
```

と入力します。Time はそれぞれの患者さんの生存期間です。Censor が 1 であるということは臨床試験が終了する前に event（死亡、再発など）が起こった場合、Censor が 0 であるということは event が起こらなかった場合を示しています。上で Censor1 をみると全ての患者さんで event が発生していることがわかります。それでは survival analysis を始めましょう。

```
. stset time Censor1
```

```
      failure event:  Censor1 ~= 0 & Censor1 ~= .
```

```
obs. time interval:  (0, time]
```

```
exit on or before:  failure
```

```
-----
33 total obs.
```

```
0 exclusions
```

33 obs. remaining, representing
33 failures in single record/single failure data
1349 total analysis time at risk, at risk from t = 0
earliest observed entry t = 0
last observed exit t = 156

まずはグラフを書いてみましょう。

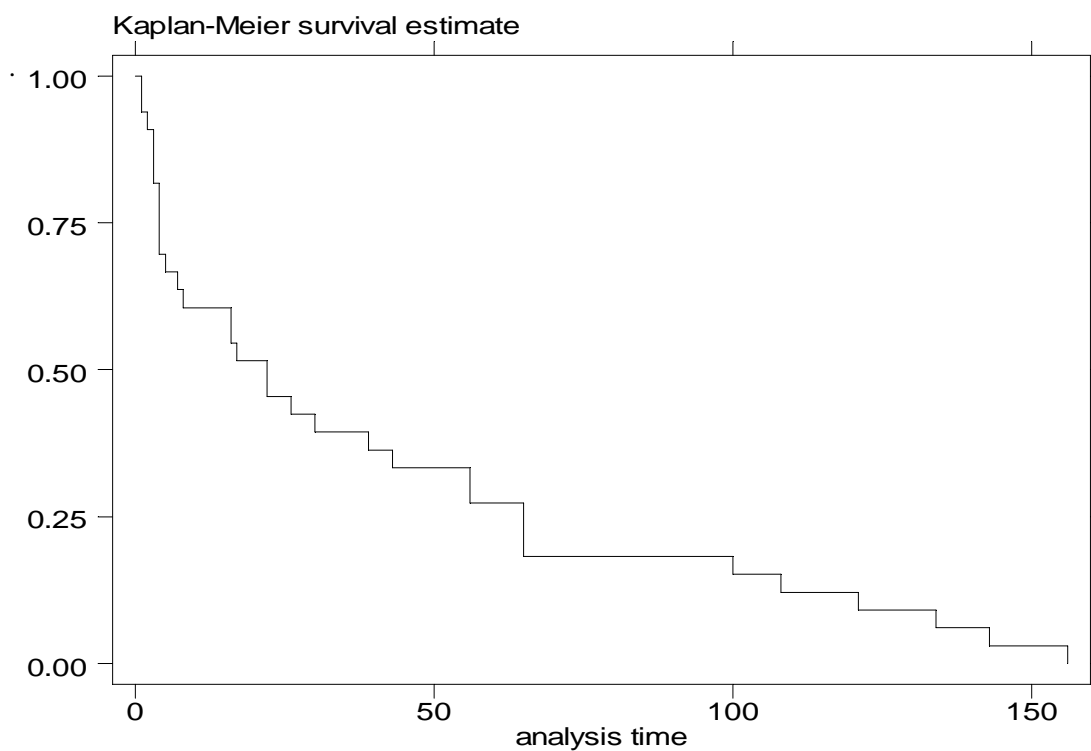
```
sts graph
```

で Kaplan-Meier survival curve を描くことができます。グラフをコピーするときは、Edit/copy/ Microsoft Word などを開き編集よりペーストを選択すれば OK です。

```
. sts graph
```

```
failure _d: Censor1
```

```
analysis time _t: time
```



この生存曲線は単に多発性骨髄腫の患者さんの状態を時間で追っただけです。

そのデータの詳細は

`sts list` で確認できます。個々の time point における survival function, standard error, 95% CI をチェックできます。

```
. sts list
```

```
failure _d: Censor1  
analysis time _t: time
```

Time	Beg. Total	Net Fail	Lost	Survivor Function	Std. Error	[95% Conf. Int.]	
1	33	2	0	0.9394	0.0415	0.7788	0.9845
2	31	1	0	0.9091	0.0500	0.7441	0.9697
3	30	3	0	0.8182	0.0671	0.6394	0.9139
4	27	4	0	0.6970	0.0800	0.5101	0.8240
5	23	1	0	0.6667	0.0821	0.4794	0.7996
7	22	1	0	0.6364	0.0837	0.4495	0.7746
8	21	1	0	0.6061	0.0851	0.4201	0.7489
16	20	2	0	0.5455	0.0867	0.3630	0.6959
17	18	1	0	0.5152	0.0870	0.3354	0.6685
22	17	2	0	0.4545	0.0867	0.2818	0.6121
26	15	1	0	0.4242	0.0860	0.2559	0.5831
30	14	1	0	0.3939	0.0851	0.2306	0.5535
39	13	1	0	0.3636	0.0837	0.2059	0.5234
43	12	1	0	0.3333	0.0821	0.1819	0.4926
56	11	2	0	0.2727	0.0775	0.1360	0.4290
65	9	3	0	0.1818	0.0671	0.0738	0.3279
100	6	1	0	0.1515	0.0624	0.0553	0.2922
108	5	1	0	0.1212	0.0568	0.0383	0.2553
121	4	1	0	0.0909	0.0500	0.0233	0.2167
134	3	1	0	0.0606	0.0415	0.0108	0.1762
143	2	1	0	0.0303	0.0298	0.0023	0.1335
156	1	1	0	0.0000	.	.	.

しかし我々の検討したいものは、予後因子などによる2群間の比較です。まず年齢によって若年者(aggroup = 1), と高齢者(aggroup = 2)に分けてみました。

まず数値を `sts list, by(aggroup)` でみてみましょう。

```
. sts list, by(aggroup)
```

```

      failure _d: Censor1
analysis time _t: time

      Beg.      Net      Survivor      Std.
Time  Total  Fail  Lost      Function  Error  [95% Conf. Int.]
-----
aggroup=1
   1     17    2    0      0.8824  0.0781  0.6060  0.9692
   4     15    1    0      0.8235  0.0925  0.5471  0.9394
   5     14    1    0      0.7647  0.1029  0.4883  0.9045
  16     13    1    0      0.7059  0.1105  0.4315  0.8656
  22     12    1    0      0.6471  0.1159  0.3771  0.8234
  26     11    1    0      0.5882  0.1194  0.3254  0.7782
  39     10    1    0      0.5294  0.1211  0.2762  0.7303
  56     9     1    0      0.4706  0.1211  0.2296  0.6797
  65     8     2    0      0.3529  0.1159  0.1448  0.5704
 100     6     1    0      0.2941  0.1105  0.1071  0.5115
 108     5     1    0      0.2353  0.1029  0.0731  0.4492
 121     4     1    0      0.1765  0.0925  0.0435  0.3830
 134     3     1    0      0.1176  0.0781  0.0196  0.3120
 143     2     1    0      0.0588  0.0571  0.0039  0.2350
 156     1     1    0      0.0000  .      .      .
aggroup=2
   2     16    1    0      0.9375  0.0605  0.6323  0.9910
   3     15    3    0      0.7500  0.1083  0.4634  0.8980
   4     12    3    0      0.5625  0.1240  0.2954  0.7622
   7     9     1    0      0.5000  0.1250  0.2452  0.7105
   8     8     1    0      0.4375  0.1240  0.1981  0.6556
  16     7     1    0      0.3750  0.1210  0.1542  0.5977
  17     6     1    0      0.3125  0.1159  0.1139  0.5365

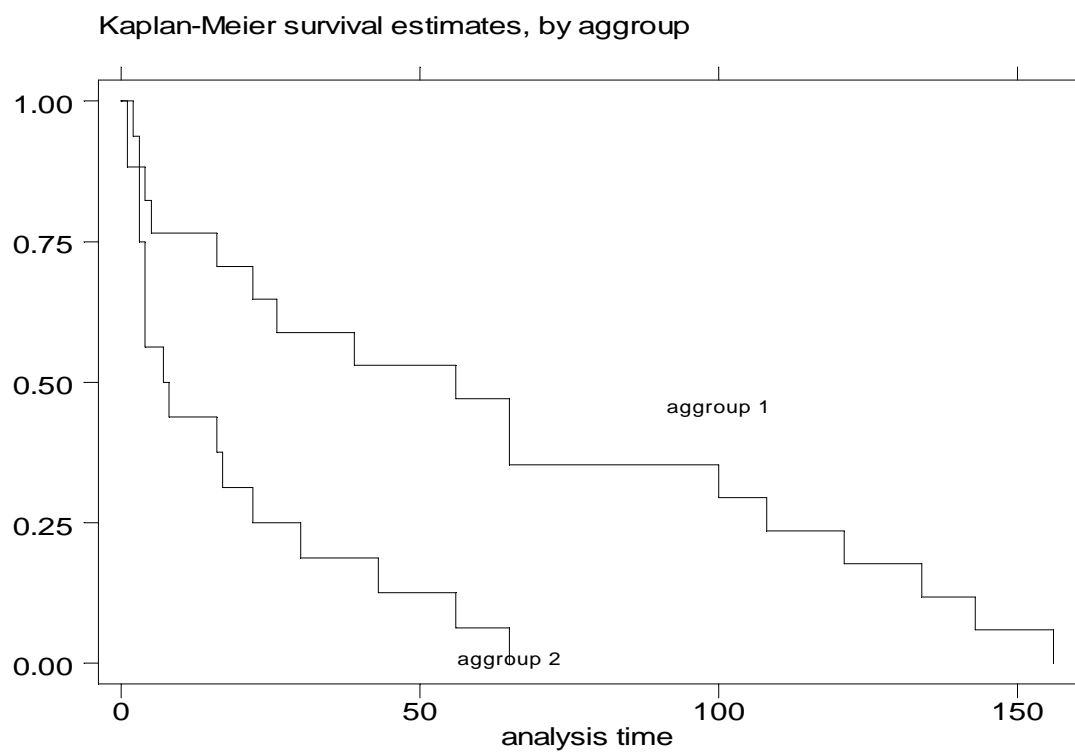
```

22	5	1	0	0.2500	0.1083	0.0775	0.4716
30	4	1	0	0.1875	0.0976	0.0460	0.4025
43	3	1	0	0.1250	0.0827	0.0207	0.3280
56	2	1	0	0.0625	0.0605	0.0041	0.2470
65	1	1	0	0.0000	.	.	.

さてこれをグラフにしてみます。

```
. sts graph, by(aggroup)
```

```
failure _d: Censor1
analysis time _t: time
```



みどころ Aggroup 1の方が予後がよさそうです。統計学的にはどうでしょうか？

`sts test aggroup` をタイプすることによって Log-rank test を行なうことができます。

```
. sts test aggroup
```

```
failure _d: Censor1  
analysis time _t: time
```

Log-rank test for equality of survivor functions

```
-----  
      | Events  
aggrou | observed   expected  
-----+-----  
1      |      17      23.70  
2      |      16      9.30  
-----+-----  
Total  |      33      33.00
```

```
chi2(1) =      8.45  
Pr>chi2 =      0.0037
```

$P = 0.0037$ で若年群と高齢群で生存曲線に差があることが統計学的に証明されました。Wilcoxon 法にても検討することができます。`sts test aggroup, wilcoxon` とタイプしてみてください。

```
. sts test aggroup, wilcoxon
```

```
failure _d: Censor1  
analysis time _t: time
```

Wilcoxon (Breslow) test for equality of survivor functions

```
-----  
      | Events  
aggrou | observed   expected   Sum of  
      |                   ranks
```



```

-----+-----
1      |      17      23.70      -128
2      |      16      9.30       128
-----+-----
Total  |      33      33.00       0

      chi2(1) =      5.60
      Pr>chi2 =      0.0180

```

次にデータをまとめてみましょう。`stsum`とタイプしてください。

```

. stsum

      failure _d: Censor1
      analysis time _t: time

      |      incidence      no. of |----- Survival time -----|
      | time at risk      rate      subjects      25%      50%      75%
-----+-----
total |      1349      .0244626      33      4      22      65

```

半数の患者さんは22ヶ月生存していることがわかります。それでは年齢別にみてみましょう。`stsum, by(aggroup)`とタイプしてください。

```

. stsum, by(aggroup)

      failure _d: Censor1
      analysis time _t: time

      |      incidence      no. of |----- Survival time -----|
aggroup | time at risk      rate      subjects      25%      50%      75%
-----+-----
      1 |      1062      .0160075      17      16      56      108
      2 |      287      .0557491      16      4      8      30
-----+-----
total |      1349      .0244626      33      4      22      65

```

若年者の半数は 56 ヶ月生存しているのに対して、高齢者では僅か 8 ヶ月しか生存していません。

次に Weibull について評価します。 `stweib wbc aggroup` とタイプしてください。

```
. stweib wbc aggroup
```

```
      failure _d:  Censor1
      analysis time _t:  time
```

Fitting constant-only model:

```
Iteration 0:  log likelihood = -62.069193
Iteration 1:  log likelihood = -60.223177
Iteration 2:  log likelihood = -60.205864
Iteration 3:  log likelihood = -60.205863
```

Fitting full model:

```
Iteration 0:  log likelihood = -60.205863
Iteration 1:  log likelihood = -56.255704
Iteration 2:  log likelihood = -54.965544
Iteration 3:  log likelihood = -54.956646
Iteration 4:  log likelihood = -54.956643
```

Weibull regression -- log relative-hazard form

```
No. of subjects =          33          Number of obs =          33
No. of failures =          33
Time at risk    =          1349

                                LR chi2(2)    =          10.50
Log likelihood = -54.956643          Prob > chi2    =          0.0053
```

```
-----
      _t | Haz. Ratio  Std. Err.      z    P>|z|    [95% Conf. Interval]
```

```

-----+-----
      wbc |   1.000007   4.77e-06   1.383   0.167   .9999972   1.000016
agggroup |   2.795969   1.085342   2.649   0.008   1.306508   5.983461
-----+-----
      /ln_p |  -.0849599   .1405221  -0.605   0.545   -.3603782   .1904583
-----+-----
      p |   .9185491   .1290765                .6974125   1.209804
      1/p |   1.088673   .1529827                .8265802   1.433872
-----+-----

```

白血球数と年齢について検討してみましたが、白血球と生存の間に関係はみいだせませんでした。Kaplan-Meyer では1つの予後因子についてしか検討することはできませんが、Cox では2つ以上について同時に検討できるメリットがあります。次に Cox Regression で評価してみます。`stcox wbc aggroup`とタイプしてみてください。

```
. stcox wbc aggroup
```

```

      failure _d: Censor1
      analysis time _t: time

Iteration 0:  log likelihood = -85.996943
Iteration 1:  log likelihood = -81.282571
Iteration 2:  log likelihood = -81.145567
Iteration 3:  log likelihood = -81.145443
Refining estimates:
Iteration 0:  log likelihood = -81.145443

Cox regression -- Breslow method for ties

No. of subjects =          33          Number of obs =          33
No. of failures =          33
Time at risk    =          1349

LR chi2(2)      =          9.70
Log likelihood = -81.145443          Prob > chi2    =          0.0078

```

```
-----+-----
      _t |
```

_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
wbc	1.000008	4.96e-06	1.540	0.124	.9999979	1.000017
agggroup	2.814298	1.182191	2.463	0.014	1.235405	6.411074

年齢は有意に生存率に関係しますが、白血球数は関係ありませんでした。

臨床試験終了時（156ヶ月）全ての患者さんは event（死亡）を持っていました。仮に何人かの患者さんはまだ生存しているのに臨床試験が終了したとします。彼らは結果がわからない宙ぶらりんの状態です。しかしこれらの患者さんの結果を捨ててしまいますか？それはもったいないので、センサーとして扱います。それでは最初に戻って Censor1 でなく、Censor2 で統計処理を行なうことをコンピュータに伝えます。

`stset time Censor2`とタイプしてください。

```
.stset time Censor2
```

```
failure event: Censor2 =0 & Censor2 =.
```

```
Obs. Time interval: (0, time)
```

```
Exit on or before: failure
```

```
33 total obs.
```

```
0 exclusions
```

```
33 obs. Remaining, representing
```

```
27 failures in single records/single failure data
```

```
1349 total analysis time at risk, at risk from t = 0
```

```
earliest observed entry t = 0
```

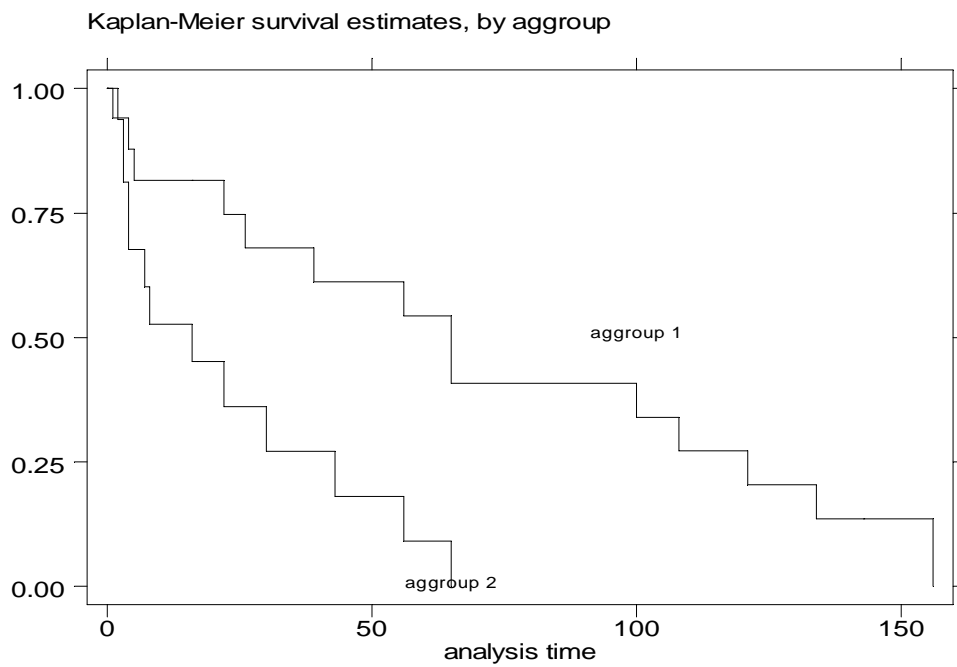
```
last observed exit t = 156
```

同様に年齢別生存曲線を描いてみます。`sts graph, by(aggroup)`とタイプしてください。

```
. sts graph, by(aggroup)
```

```
failure _d: Cencor2
```

```
analysis time _t: time
```

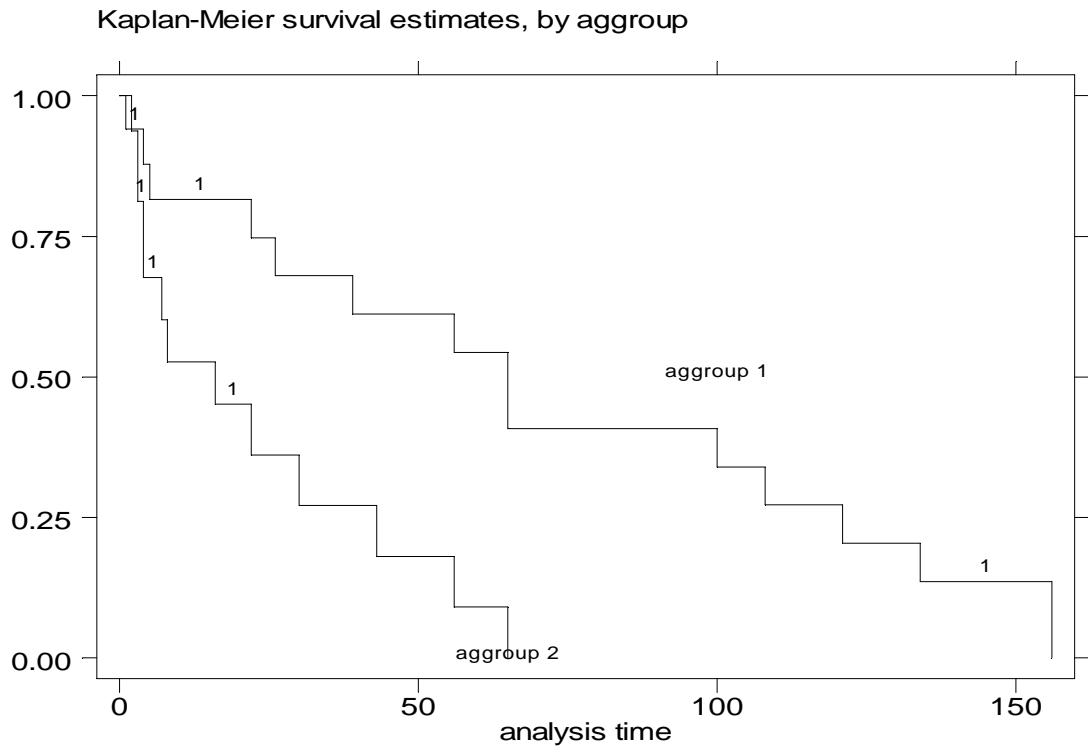


しかしこの生存曲線ではどこで患者さんがセンサーとなったか、例えば相当長期に生存していたのか、短い時間なのかわかりません。次にグラフ上にセンサーに標を入れます。

```
. sts graph, by(aggroup) lost
```

```
failure _d: Cencor2
```

```
analysis time _t: time
```



1はセンサーを示しています。以下は Censor2 において Censor1 と同様に統計学的処理を行なったものです。

```
. sts test aggroup
```

```
failure _d: Cencor2  
analysis time _t: time
```

Log-rank test for equality of survivor functions

```
-----  
      | Events  
aggrou | observed    expected  
-----+-----  
1      |      14      19.82  
2      |      13       7.18  
-----+-----  
Total  |      27      27.00
```

```
chi2(1) =      7.95  
Pr>chi2 =      0.0048
```

```
. sts test aggroup, wilcoxon
```

```
failure _d: Cencor2  
analysis time _t: time
```

Wilcoxon (Breslow) test for equality of survivor functions

```
-----  
      | Events                Sum of  
aggrou | observed    expected    ranks  
-----+-----  
1      |      14      19.82      -109
```

```
2 | 13 7.18 109
```

```
-----+-----
```

```
Total | 27 27.00 0
```

```
chi2(1) = 5.43
```

```
Pr>chi2 = 0.0198
```

```
. stsum
```

```
failure _d: Cencor2
```

```
analysis time _t: time
```

```
 | incidence no. of |----- Survival time -----|  
 | time at risk rate subjects 25% 50% 75%  
-----+-----  
total | 1349 .0200148 33 5 30 65
```

```
. stweib wbc aggroup
```

```
failure _d: Cencor2
```

```
analysis time _t: time
```

```
Fitting constant-only model:
```

```
Iteration 0: log likelihood = -52.277187
```

```
Iteration 1: log likelihood = -51.520396
```

```
Iteration 2: log likelihood = -51.516734
```

```
Iteration 3: log likelihood = -51.516734
```

```
Fitting full model:
```

```
Iteration 0: log likelihood = -51.516734
```

```
Iteration 1: log likelihood = -48.218707
```

```
Iteration 2: log likelihood = -46.602149
```

```
Iteration 3: log likelihood = -46.584458
```


Iteration 4: log likelihood = -46.584451

Iteration 5: log likelihood = -46.584451

Weibull regression -- log relative-hazard form

No. of subjects = 33 Number of obs = 33

No. of failures = 27

Time at risk = 1349

LR chi2(2) = 9.86

Log likelihood = -46.584451 Prob > chi2 = 0.0072

```
-----
```

	_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
	wbc	1.000007	5.21e-06	1.418	0.156	.9999972	1.000018
	agggroup	2.975792	1.289802	2.516	0.012	1.272521	6.958893
	/ln_p	.0041808	.1539206	0.027	0.978	-.2974981	.3058597
	p	1.00419	.1545655			.742674	1.357792
	1/p	.9958279	.1532785			.7364899	1.346486

```
-----
```

. stcox wbc agggroup

failure _d: Cencor2

analysis time _t: time

Iteration 0: log likelihood = -69.22415

Iteration 1: log likelihood = -64.723793

Iteration 2: log likelihood = -64.557569

Iteration 3: log likelihood = -64.557406

Refining estimates:

Iteration 0: log likelihood = -64.557406

Cox regression -- Breslow method for ties

```

No. of subjects =      33              Number of obs =      33
No. of failures =      27
Time at risk   =      1349

LR chi2(2)      =      9.33
Log likelihood = -64.557406          Prob > chi2    =      0.0094

```

```

-----
      _t |
      _d | Haz. Ratio  Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      wbc |  1.000009  5.42e-06    1.571  0.116    .9999979  1.000019
      aggroup |  3.063232  1.434787    2.390  0.017    1.223166  7.671392
-----

```

Cox 法においても年齢は予後に関係するが、白血球数は関係ないという結果となりました。